GEOSTATISTICAL UNCERTAINTY ANALYSIS IN SEDIMENT GRAIN SIZE MAPPING WITH HIGH-RESOLUTION REMOTE SENSING IMAGERY

No-Wook Park and Kwang-Hoon Chi

Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources
30 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Korea
E-mail: nwpark@kigam.re.kr

ABSTRACT This paper presents a geostatistical methodology to model local uncertainty in spatial estimation of sediment grain size with high-resolution remote sensing imagery. Within a multi-Gaussian framework, the IKONOS imagery is used as local means both to estimate the grain size values and to model local uncertainty at unsample locations. A conditional cumulative distribution function (ccdf) at any locations is defined by mean and variance values which can be estimated by multi-Gaussian kriging with local means. Two ccdf statistics including condition variance and interquartile range are used here as measures of local uncertainty and are compared through a cross validation analysis. In addition to local uncertainty measures, the probabilities of not exceeding or exceeding any grain size value at any locations are retrieved and mapped from the local ccdf models. A case study of Baramarae beach, Korea is carried out to illustrate the potential of geostatistical uncertainty modeling.

KEY WORDS: Uncertainty, Multi-Gaussian Kriging, ccdf, IKONOS

1. INTRODUCTION

Traditionally, environmental thematic mapping such as grain size distribution and sedimentary facies maps are based on field observation. Since field survey are limited by the cost of sampling and the short exposure time above water of tidal flats, sparse observation data are generally available. Thematic mapping or further analysis requires exhaustive attribute values for the entire study area. Thus interpolation procedures are commonly adopted in order to obtain attribute values at unsampled locations. Among various interpolation algorithms, geostatistical kriging can account for spatial variability in the interpolation procedure. If any secondary information can be obtained that has a reasonable correlation with the primary attribute of interest and is exhaustively sampled, the estimation error can be reduced when compared with the case using only sparsely sampled data (Goovaerts, 2000).

If remote sensing imagery is considered as an information source that provides exhaustive information over the area of interest, it can be incorporated directly into geostatistical kriging for spatial estimation. Park et al. (2006) applied practical multivariate geostatisical algorithms to spatial estimation of sediment grain size by integrating high-resolution IKONOS imagery with sparsely sampled grain size data. The case study results indicate that accounting for the IKONOS imagery via simple kriging with local means could reflect detailed surface characteristics with less smoothing effects. Cross validation results also showed that the multivariate geostatistical algorithms produced the better prediction capabilities than traditional univariate geostatistical kriging algorithms.

Although the secondary information can be a useful information source for the geostatistical mapping procedure, however, the uncertainty attached to the estimation should be assessed for further decision-making or analysis.

This paper describes the potential of geostatisical kriging for uncertainty modeling in spatial mapping of sediment grain size with high-resolution remote sensing imagery. In this study, local uncertainty is modelled by a multi-Gaussian model as the parametric approach and then the uncertainty regarding grain size distribution is interpreted. Multi-Gaussian kriging with varying local means is adopted both to model the uncertainty about grain size for all pixels and to integrate high-resolution remote sensing imagery. We illustrate the presented schemes by undertaking a case study of Baramarae beach, Korea

2. STUDY AREA AND DATA

The present study was conducted at Baramare beach, located in the south of Anmyeondo, Korea. In the study area, sea stacks including the Halmi and Seomot isles with wave-cut platform and sand dune now block the swell and a large tidal flat has developed (Jang et al., 2003).

High-resolution IKONOS imagery acquired on February 26, 2001 was used as the exhaustive information source for grain size mapping. In consideration of the season and tide condition at the time of the IKONOS imagery acquisition, we conducted a field survey on February 26, 2002. We collected sediment sampled at 53 points. Sample fractions were obtained at the laboratory using a set of sieves arranged

with a half-phi interval and the mean grain sizes were obtained in phi units.

3. GEOTATISTICAL METHODOLOGY

3.1 Multi-Gaussian Kriging

When considering the local uncertainty assessment for spatial estimation in geostatistics, it is usually modelled by the conditional cumulative distribution function (ccdf) (Goovaerts, 1997). In this study, a multi-Gaussian model is adopted to model ccdfs.

Our final goals for this geostatistical analysis are to estimate the grain size value $z(\mathbf{u})$ and to model the ccdf at an unsampled location $\mathbf{u}$ using the grain size values at n sample locations $\{ z(\mathbf{u}_\alpha), \alpha=1,\cdots,n \}$. In order to apply multi-Gaussian kriging, an intrinsic assumption is that the grain size values follow a standard Gaussian distribution. To satisfy this assumption, a normal score transform is generally applied to the original data values (Deutsch and Journel, 1998). Once a new transformed variable $y(\mathbf{u}_\alpha)$ has been obtained, the normal score value $y^*_{SK}(\mathbf{u})$ at any location $\mathbf{u}$ is the simple kriging estimate as linear combination of the surrounding normal score transformed grain size data at sample location $\mathbf{u}_\alpha$.

Under the multi-Gaussian framework, the mean and variance, which are two parameters of the ccdf $F(\mathbf{u};y|(\text{info}))$ at any location $\mathbf{u}$, correspond to simple kriging estimate $y^*_{SK}(\mathbf{u})$ and variance $\sigma^{2*}_{SK}(\mathbf{u})$ at $\mathbf{u}$, respectively, and then the ccdf model is defined as (Goovaerts, 1997):

$$F(\mathbf{u};y \mid (\text{info})) = G[(y - y^*_{SK}(\mathbf{u}))/\sigma^*_{SK}(\mathbf{u})] \qquad (1)$$

where (info) and G are the nearby sampled data and the standard Gaussian cumulative distribution function, respectively.

Once the ccdf modeling has been done in the Gaussian space, the back-transform of the results is carried out to obtain values in the original space. By following Saito and Goovaerts (200), the kriging estimation $z^*_{MG}(\mathbf{u})$ is empirically obtained as the E-type estimaton of the z ccdf.

3.2 Integration of IKONOS Imagery

In this study, IKONOS imagery was incorporated into multi-Gaussian kriging to derive the local means or trend of the normal score transformed grain size values. Among various multivariate kriging algorithms for the integration of exhaustively sampled secondary information, Simple kriging with local means (SKlm) was adopted that can incorporate several secondary information sources by a linear or nonlinear calibration procedure.

In the Gaussian space, the SKlm estimate of the normal score value $y^*_{SKlm}(\mathbf{u})$, which is also a mean value of the y ccdf, is defined as:

$$y^*_{SKlm}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha [y(\mathbf{u}_\alpha) - m^*_y(\mathbf{u}_\alpha)] + m^*_y(\mathbf{u}_\alpha) \qquad (2)$$

Because the SKlm approach is the kriging of residuals, the residual values at any location $\mathbf{u}$ are first estimated in SKlm using neighboring residual values at sample location $\mathbf{u}_\alpha$ and then they are added to the local means $m^*_y(\mathbf{u})$. The sum of the SKlm variance and the square of standard error of the local mean estimate is used as the variance of the y ccdf.

In this study, the local means were estimated by applying a generalized additive model (GAM). The GAM extends a generalized linear model by nonlinear and smoothing operations (Hastie and Tibshirani, 1990).

4. RESULTS

4.1 Derivation of Local Means

Prior to geostatistical analysis, we first examined how strong each spectral band of IKONOS imagery was correlated with grain size at the sample locations. The normal score transformed values show a reasonable linear correlation relationship with the IKONOS imagery. The highest correlation was observed in the green band with a value of -0.859. Next was the red band with -0.849.

Based on this correlation, local means were derived from IKONOS imagery using GAM. Cubic B-spline functions were used as nonparametric smoothing operators. The use of all spectral bands showed the smallest residual deviance among various band combinations, and thus local means derived from all the bands were finally used.

4.2 Variogram Analysis Results

To examine the spatial variability of grain size distribution, we computed the variogram map and experimental variograms of normal score transformed grain size values.

From the variogram map (not shown here), a strong anisotropy was observed, orientated NW-SE (i.e., 135° counter-clockwise from the EW direction). The reason why reflectance values in the NW–SE direction are more uniform than that in the NE-SW direction may be explained by topography in the study area. Halmi and Seomot isles lie linearly NW-SE at the southern end of the Baramarae Beach and inhibit tidal currents and ocean swell entering from the open sea. As a result, tides may not flow in from the NE–SW direction but from the NW–SE direction. Furthermore, an analysis of wind ripples in the study area reveals that the seasonal northwest monsoon causes fine-grained sediment to move NW–SE.

Based on the anisotropy, we computed experimental variograms in the major NW-SE direction and the perpendicular NE-SW trend with an angular tolerance of 22.5° and then fitted them using the geometric anisotropy model. The variogram of the residual values of normal score transforms were spatially correlated because of

their bounded spatial patterns with very low nugget effects.

4.3 Geostatistical Mapping Results

Spatial estimation results of kriging variants are presented in Figure 1. A strong NW-SE continuity was observed in the univariate multi-Gaussian kriging result. This resulted from the geometric anisotropy variogram model with a strong anisotropy in the major NW-SE direction. The smoothing patterns that are typical characteristics of kriging were also observed. The sand spit and beach in front of Halmi isle were clearly presented in the result.

In the kriging result with IKONOS imagery, the local spatial grain size patterns were reflected well with less smoothing effects. This indicates that IKONOS imagery significantly influenced the kriging results and allows a better interpretation of the local details of grain size.

To investigate which kriging algorithm generated the most accurate estimates of grain size, a cross validation based on the leave-one-out approach was carried out and the estimation capabilities were quantitatively evaluated. As expected, the incorporation of IKONOS imagery leads to smaller prediction errors. The mean square error values of multi-Gaussian kriging and SKlm with GAM were 0.231 and 0.093, respectively.

When compared with univariate kriging (i.e., multi-Gaussian kriging), improvements of about 60% were observed in SKlm with GAM estimates. These quantitative evaluation results confirm that accounting for secondary information that is well correlated with primary attributes can contribute to the improvement of prediction capability at unsampled locations.

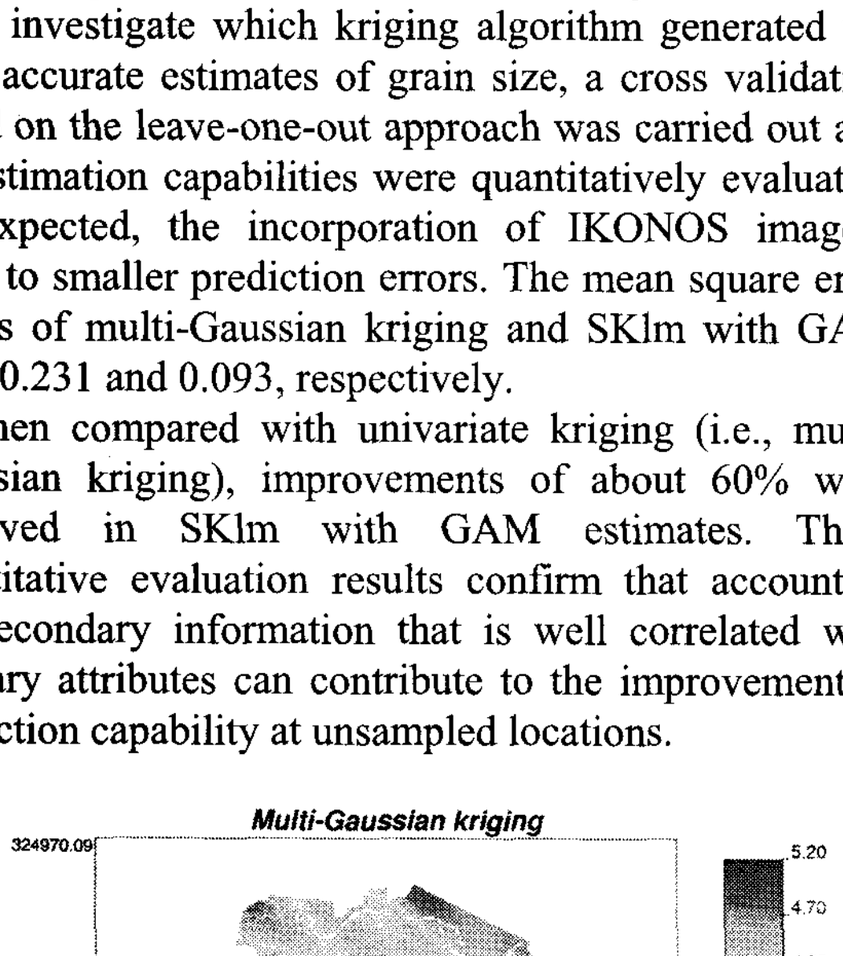

Figure 1. Grain size distributions generated by multi-Gaussian kriging (top) and SKlm with GAM (bottom).

4.4 Local Uncertainty Analysis Results

The uncertainty attached to the grain size estimation can be easily assessed since the ccdf model at any location is already fully known from kriging analysis. Two ccdf statistics including conditional variance and interquartile range are used here as measures of local uncertainty (Goovaerts, 1997) and are compared with square errors of the cross validation analysis in Table 1. If the spread measures of conditional probability density function are larger, the uncertainty attached to prediction is greater.

Mean and standard deviation values of two local uncertainty measures for SKlm with GAM were significantly smaller than those of multi-Gaussian kriging. A very strong correlation between square error and local uncertainty was not observed due to very low square error values of SKlm with GAM. However, stronger correlation with the square error could be observed for SKlm with GAM, which matches well with our intuition that the larger the uncertainty value is, the larger the prediction error will be.

Table 1. Summary statistics for local uncertainty measures

Statistics		Multi-Gaussian kriging	SKlm with GAM
Standard dev. of MSE		0.308	0.119
Conditional variance	Mean	0.195	0.141
	Standard dev.	0.128	0.085
	Correlation with squared errors	0.420	0.450
	Rank correlation with squared errors	0.380	0.546
Interquartile range	Mean	0.549	0.483
	Standard dev.	0.306	0.265
	Correlation with squared errors	0.339	0.458
	Rank correlation with squared errors	0.420	0.552

In addition to local uncertainty measures based on cross validation, the probabilities of not exceeding or exceeding any grain size value at any location were retrieved from the local ccdf models. By following Wentworth classification criteria, the thresholding value of 4 phi (0.0625 mm) between sand and silt classes was selected. If the grain size in phi units is smaller than 4 phi and larger than –1 phi, the grain is classified as being in the sand category. The probability of belonging to the sand class was computed directly from the normal score transform of the target threshold value of 4 phi and the mean and variance values of the y ccdf. Conversely, the probability of belonging to the silt class is easily retrieved by subtracting the probability of belonging to the sand class from one.

The probability of belonging to the sand class for kriging algorithms applied in this study is shown in Figure 2. By observing the areas that have more than 80% probability of belonging to sand in multi-Gaussian kriging, it seems likely that it was relatively overevaluated compared to the field data. Although it is

possible to analyze the regular direction of sand in existing N-S directions in multi-Gaussian kriging because there is a high probability of sand occurring in the entire tidal flat excluding the bay-mouse area, the results show a considerable difference from the actual field observations. On the other hand, SKlm with GAM shows very similar results to the field observations in regions that have more than 80% probability of belong to sand. In other words, SKlm with GAM reflects the field data more accurately than multi-Gaussian kriging in the expression of patterns and detail. In particular, the probability of belonging to sand in the bay-mouse areas and tidal channels is shown to be relatively high, which was not found in the result by multi-Gaussian kriging.

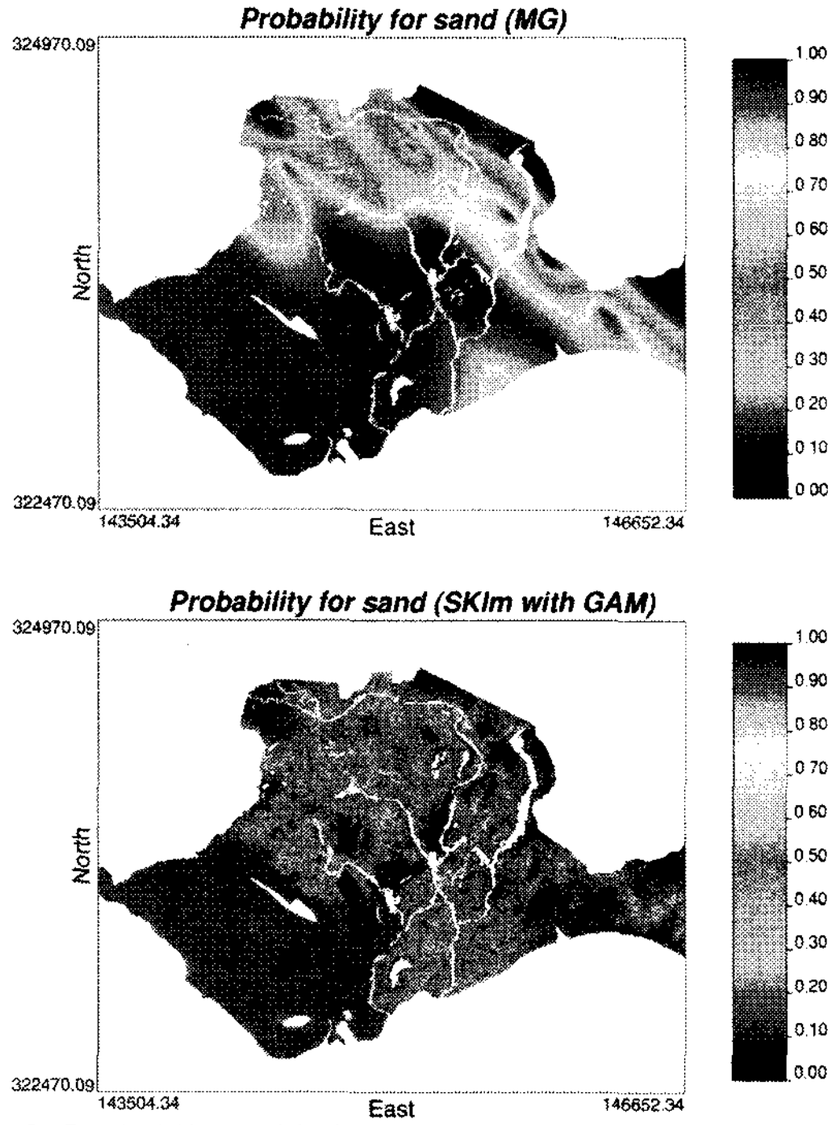

Figure 2. Probability of belonging to the sand class for kriging algorithms applied in this study.

5. CONCLUSIONS

This paper has demonstrated the geostatistical anaylsis procedures for spatial estimation and local uncertainty modeling with high-resolution remote sensing imagery. Especially, the main focus was on how multivariate geostatistical kriging (in this study, SKlm within a multi-Gaussian framework) can be used to map not only detailed surface distribution of sediment grain size but also the uncertainty attached to the estimation.

Experimental results from a case study of Baramarae beach, Korea showed that accounting for high-resolution IKONOS imagery via SKlm with GAM, despite its poor spectral resolution, greatly improved the prediction accuracy and reflected the detailed local variation of grain size distribution in the study area, as compared with traditional univarite multi-Gaussian kriging. When the result of spatial estimation is used as input for any environmental models, uncertainty or error propagation is one of critical issues. Especially, as the result of local uncertainty modeling, the spatial mapping of probability of belonging to certain class can be used preliminary information to evaluate the reliability of spatial estimation results or to analyze the effects of uncertainty propagation.

From a methodological point of view, the uncertainty model adopted in this paper is a parametric multi-Gaussian model. As for other uncertainty model, a nonparametric approach such as indicator kriging can be applied. Since what type of uncertainty model is adopted is important, comparative study should be undertaken. In addition to local uncertainty modeling, the uncertainty about any attribute values to be assessed at many locations simultaneously, namely spatial uncertainty, should be evaluated by stochastic simulation for a certain application purpose.

6. REFERENCES

Deutsch, C.V. and Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228, pp.113-129.

Jang, D.-H., Kim, M.-K., and Kim, K.-H., 2003. Change of the sedimentary environment system at the Baramarae beach, Anmyeondo area in west coast of Korea. *Journal of the Korean Geomophological Association*, 10, pp.143-155.

Park, N.-W., Chi, K.-H., and Jang, D.-H., 2006. Geostatistical integration of high-resolution remote sensing data in spatial estimation of grain size. In: ISRS 2006 PORSEC, Busan, Korea, CD-rom publication.

7. CONCLUSIONS

This work was supported by the Korea Aerospace Research Institute, Korea. We thank Prof. D.-H. Jang from Kongju National University, Korea for providing the sampling data used in this study.